

Araştırma Makalesi

İKİNCİ EL ARAÇ DEĞERLEMESİNDE MAKİNE ÖĞRENME TEKNİKLERİNİN UYGULANMASI

Esra DERE[†], Mustafa Cem KASAPBAŞI^{††}[†] İstanbul Ticaret Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Ana Bilim Dalı, İstanbul, Türkiye^{††} İstanbul Ticaret Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Ana Bilim Dalı, İstanbul, Türkiyeesra.dere@istanbulticaret.edu.tr, mckasapbasi@ticaret.edu.tr

0000-0003-1598-5853, 0000-0001-6444-6659

Atıf/Citation: ESRA, D., KASAPBAŞI, M. C., (2023). İkinci El Araç Değerlemesinde Makine Öğrenme Tekniklerinin Uygulanması, Journal of Technology and Applied Sciences 6(2), s.37-51 , DOI: 10.56809/icujtas.1329496

ÖZ

Son yıllarda ikinci el araçların pazar hacmi artmıştır. Bu pazarda satıcı ve alıcı için doğru fiyatlandırma oldukça önemlidir. Son kullanıcı veya kurumlar için ikinci el araç değerlemesine ya da kiralanmasında yardımcı olacak sistemsel bir yapıya ihtiyaç bulunmaktadır. Bu çalışmada ilgili ikinci el araç ilanların yer aldığı sitelerden Selenium test aracı ile 26.000 üzerinde veri toplanmış ve bu veriler üzerinde veri ön işleme (temizleme, dönüştürme vs.) adımları uygulanmıştır. Makine öğrenme teknikleri KNIME Analytics Platform veri madenciliği programının 4.7.3 sürümünde uygulanarak ikinci el araç fiyatı tahmin edilmeye çalışılmış ve sonuçlar birbiriyle karşılaştırılmıştır. Performans ölçülürken R^2 kullanılmıştır. Sonuçlar değerlendirildiğinde Lineer regresyon 0,56 R^2 , Random Forest 0,83 R^2 , GBoosted 0,81 R^2 ve Tree Ensemble 0,82 R^2 oranıyla tahminleme için başarılı sonuçlar elde edilmiştir.

Anahtar Kelimeler: Makine Öğrenmesi, KNIME, İkinci El Araç Fiyat Tahmini

APPLICATION OF MACHINE LEARNING TECHNIQUES IN USED VEHICLE VALUATION

ABSTRACT

The market volume of second-hand vehicles has increased in recent years. In this market, accurate pricing is very important for the seller and the buyer. There is a need for a systematic structure that will help end-users or organizations in the valuation or leasing of used vehicles. In this study, over 26,000 data were collected from the relevant used car classifieds websites with Selenium test tool and data preprocessing (cleaning, transformation, etc.) steps were applied on these data. Machine learning techniques were applied in version 4.7.3 of the KNIME Analytics Platform data mining program to predict the used car price and the results were compared with each other. R^2 was used to measure performance. When the results are evaluated, successful results were obtained with Linear regression 0.56 R^2 , Random Forest 0.83 R^2 , GBoosted 0.81 R^2 and Tree Ensemble 0.82 R^2 .

Keywords: Machine Learning, KNIME, Second Hand Car Price Estimate

Geliş/Received : 18.07.2023
Gözden Geçirme/Revised : 21.07.2023
Kabul/Accepted : 28.07.2023

1. Giriş

Günümüzde yüksek ÖTV oranları, bayilerde veya galerilerde araç bulunamaması gibi nedenlerden dolayı ikinci el araç satış hacmi Türkiye de ve dünyada artmıştır. Yeni otomobil satış hacmi, ikinci el satış hacminin gerisinde kalmıştır. Bu durum, ikinci el araç piyasasının gün geçtikçe daha da önem kazandığını ortaya koymaktadır. Dünyadan örnek verilecek olursa, ABD ekonomisinde ikinci el araç pazarı, sıfır araç pazar hacminden iki kat daha fazla ve bu fazlalık süreklilik göstermektedir [1]. İkinci el araç pazarları, tüketicilere maksimum kalitede uygun fiyatlı araç bulma imkanı sağlamaktadır. Bu noktada aracın gerçek değerini belirlemek hem satıcı hem de alıcı için oldukça önemlidir.

Mevcut sistemde satıcı aracının gerçek değeri konusunda yani kaçta satması gerektiğini, alıcının da aracın ederinin ne olduğu konusunda bilgi sahibi değildir. Aracın fiyatın belirlenmesinde birçok faktör bulunmaktadır. Aracın yakıt türü, vites tipi, markası, modeli, hasar durumu, yaşı ve kaporta durumu gibi özellikleri bu faktörlerden bazılarıdır. Otomotiv sektörü, ekonomide yeri oldukça önemli bir konumda olan bir sektördür. Otomobil, motorlu araç üretim sektöründe 70 gibi büyük bir orana sahiptir. Bu bilgiler ışığında otomobil endüstrisinin, dünya ekonomisindeki payının büyüklüğü anlaşılabilir [2].

Son yıllarda teknolojinin gelişmesiyle beraber veri kolayca toplanabilmektedir. Bu sayede veri depolama sistemlerinde geçmişe göre çok daha fazla veri tutulmaktadır ve depolanmaktadır. Veri madenciliği sayesinde tutulan verilerden anlamlı bilgiler çıkarılmakta ve çıkarılan bu bilgiler sayesinde karar alma süreçlerinde daha verimli hareket edilmektedir [3].

Bu çalışmada, Doğrusal (Linear) regresyon, Rastgele Orman (Random Forest) regresyon, Ağaç Topluluğu (Tree Ensemble) ve Gradyan Artırıcı (GBoosted) makine öğrenme algoritmaları kullanılarak KNIME aracı ile ikinci el araç ilanlarından yararlanılarak aracın fiyatı tahmin edilmiştir. İkinci el araç verileri Türkiye'deki ikinci el araç ilanlarının yayınlandığı siteden temin edilmiştir. Ham veri seti veri temizleme, dönüştürme gibi veri ön işleme adımlarından geçmiştir. Veri ön işleme adımından sonra veriler kendi arasında analiz edilmiş ve paylaşılmıştır. Elde edilen sonuçlar karşılaştırılmış ve ikinci el araç fiyatını belirlemedeki en doğru fiyat tahmini veren algoritma belirlenmiştir.

Makalenin birinci bölümünde ilgili çalışmaların araştırılmasına yer verilmiştir. İkinci bölümde kullanılan materyal ve metodlar hakkında bilgiler verilmiştir ve veri seti tanıtılmıştır. Üçüncü bölüm olan uygulama bölümünde ise, veri ön işleme adımları ve elde edilen analiz sonuçları, KNIME ortamında uygulamanın adımları detaylı olarak anlatılmıştır. Dördüncü bölümde, KNIME aracı ile uygulanan makine öğrenmesi algoritmalarının gerçekleştirilmesi ve sonuçlarına yer verilmiştir.

2. İlgili Çalışmaların Araştırılması

Bu kısımda, 2000-2020 yılları arasındaki ilgili çalışmalar araştırılmıştır.

Pazarlioglu ve Gunes (2000), çalışmalarında hedonik fiyat modeli oluşturup üzerine tartışmışlardır. Deneysel analiz sonuçlarını paylaşmışlardır. Bulanık hedonik model ve normal model tahminleri karşılaştırılmıştır. Yüksek düzeyde bilgi sağlayan en iyi bilgi birleştirme yöntemi belirlenmiştir [4].

Asilkan ve Irmak (2009), ikinci el araç fiyatının güncel ve gelecekteki fiyatını tahmin edebilmek için makine öğrenmesi yöntemlerinden Yapay Sinir ağları ile çalışmışlardır. Verileri ikinci el araç satış sitesinden elde etmişlerdir. YSA sonuçlarını zaman serisi analizleri ile elde ettikleri sonuçlarla kıyaslamışlardır. Çalışma sonucunda Yapay Sinir Ağları'nın tahminleme de başarılı ve güvenilir sonuçlar verdiği gözlenmiştir. YSA'nın otomobil sektöründe de kullanımını tavsiye etmişlerdir [5].

Namlı ve arkadaşları (2018) çalışmalarında, ikinci el araçlarda ikinci el araç satışlarında uygun bir fiyatlandırma yönteminin belirlenip belirlemeyeceğini araştırmışlardır. İstatiksel yöntemlerden olan Doğrusal Regresyon ve makine öğrenme tekniklerinden Destek Vektör Makineleri, Yapay Sinir Ağları kullanmışlardır. Algoritmalar, doğruluğunu kanıtlamak amaçlı çapraz doğrulama ve veri bölme yöntemi ile çalışmışlardır. En yüksek başarıyı DVM algoritması ile elde etmişlerdir [6].

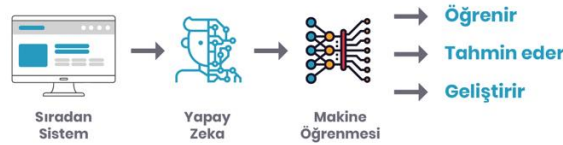
Pal ve arkadaşları (2018) çalışmalarında, ikinci el araba satış hacminin artmasıyla birlikte araç özelliklerine ve diğer araç fiyatlarına göre fiyat ataması yapabilmeyi amaçlamışlardır. Makine öğrenmesi gözetimli öğrenme yöntemlerinden olan Rastgele Orman (Random Forest) kullandılar. Eğitim için 500 karar ağacı oluşturdular. Eğitim doğruluğunu 95,82 ve test doğruluğunu 83,63 olarak elde ettiler [7].

Çelik ve Osmanoğlu (2019), ikinci el araç pazarında hızlı satış yapabilmek, doğru fiyatlandırma yapma amaçlı ideal fiyatı makine öğrenme yöntemleri ile yakalamaya çalışmışlardır. İkinci el otomobile ait 5041 adet veri setindeki 78 değişkenden 23'ü belirleyerek model oluşturdu ve makine öğrenmesi tahminleme modelinin başarısının oranını 81,15 olarak elde ettiler. Daha fazla özneliğe sahip veri seti ile daha iyi tahmin oranlarının ortaya çıkabileceğini savunmuşlardır [8].

3. Kullanılan Materyal ve Metotlar

3.1. Makine Öğrenmesi

Yapay zekanın alt alanı olan makine öğrenmesi son yıllarda artan veri miktarıyla birlikte birçok alanda uygulanmaktadır ve başarılı sonuçlar elde edilmektedir. Makine öğrenmesi geçmiş veri deneyimlerini kullanarak öğrenme ve geliştirme odaklı yeni veriye en uygun modeli istatistiksel ve algoritmalar ile bulmaya çalışır. Temel çalışma prensibi, bilinen veri üzerinden bilinmeyen veri üzerinde çıkarım yapma üzerinedir. Makine öğrenmesinin çalışma mantığı Şekil 1'de görüldüğü gibidir. Makine öğrenmesindeki amaç, veri ve deneyimden faydalanarak bilgisayara insan beynine benzer bir öğrenme yeteneği kazandırmaktır. Sınıflandırma, kümeleme ve tahmin gibi problem çeşidine göre makine öğrenme yöntemleri vardır.



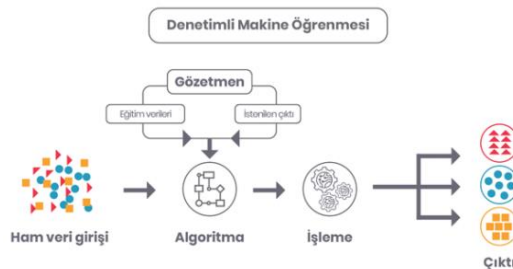
Şekil 1. Makine öğrenmesi

Makine öğrenmesi algoritma türleri;

- Denetimli (Supervised) Öğrenme,
 - Regresyon,
 - Sınıflandırma,
- Denetimsiz (Unsupervised) Öğrenme,
 - Kümeleme,
 - Boyut Azaltma
- Takviyeli – Pekiştirmeli (Reinforcement) Öğrenme

olmak üzere üç gruba ayrılır.

Gözetimli öğrenme, girdi ve çıktının arasındaki ilişkiyi öğrenmek amaçlı etiketli verilerle çalışır. Eğitim ve test kümesi oluşturularak eğitim gerçekleştirir. Şekil 2'de gözetimli öğrenmenin işleyişi görüldüğü gibidir. Sınıflandırma ve regresyon problemleri için kullanılır. En yoğun kullanılan gözetimli öğrenme algoritmaları Destek Vektör Makineleri, Karar ağaçları ve Doğrusal Regresyondur.



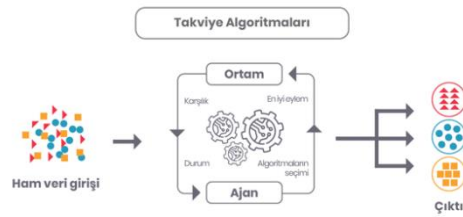
Şekil 2. Denetimli makine öğrenmesi

Gözetimsiz öğrenme, etiketsiz veri üzerinden algoritmalar kullanarak bağımsız olarak yapıları tahmin etmeye çalışır. Gözetimsiz öğrenme, veri setindeki verilerin arasındaki benzerliği veya desenleri keşfeder. Modelleri kümeler, ilişkilendirir ve boyut azaltma gibi işlevleri vardır. Şekil 3'te gözetimsiz öğrenmenin işleyişi görüldüğü gibidir.



Şekil 3. Denetimsiz makine öğrenmesi

Takviyeli öğrenme ise, oyunlarda, kontrol sistemlerinde veya robotiklerde karar verme veya karmaşık problemlerin çözümünde hangi eylemin en iyi sonuç verdiğini deneme yanılma yöntemini kullanarak çözmeye çalışır. Belirli bir ortamda hedefi en iyi şekilde gerçekleştirmeyi öğrenmeyi hedefler. Bir ajan vardır ve bu ajan bir durum karşısında aksiyon seçer ve belirli olan ortamdan buna karşılık ödül ya da ceza alır, yeni aksiyona geçer. Bu adımları tekrarlayarak en iyi aksiyonları seçmeyi öğrenir. Şekil 4'te takviyeli öğrenme işleyişi görülebilir.



Şekil 4. Takviyeli makine öğrenmesi

3.2. Makine Öğrenmesi Algoritmaları

Problem yaklaşımlarına göre Naive Bayes, M5, K-Means, Doğrusal (Linear) Regresyon, C4.5 ve Rastgele Orman (Random Forest) gibi birçok makine öğrenmesi algoritmaları bulunmaktadır.

Bu çalışmada Doğrusal (Lineer) Regresyon, Rastgele Orman (Random Forest) algoritması, Gradyan Artırıcı (Gradient Boosting) ve Ağaç Topluluğu (Tree Ensemble) algoritmaları denenmiştir. Veri, 20% doğrulama, 80% eğitim verisi olacak şekilde modelin başarısını ölçmek için ayrılmıştır. Sonuçlar makalenin sonuç kısmında verilmiştir.

Model başarısını ölçmek için farklı parametreler bulunmaktadır. Bu çalışmada R^2 değeri, MAE (Ortalama Mutlak Hata) ve RMSE (Ortalama Karese Hatanın Karekökü) parametreleri ile model performansı değerlendirilmiştir. R^2 değeri, 0,8'den büyük ise model uyumu güçlü, 0,5'ten küçük ise zayıf olarak değerlendirilmektedir. Yani modelimiz ne kadar 1'e yakın ise model başarısı o kadar yüksektir.

3.2.1. Doğrusal (Linear) Regresyon

Bağımlı değişken ile bir veya birden fazla bağımsız değişken arasında bir desen kurmayı amaçlar. Bilinen veriyle bilinmeyen veriyi analiz eder.

3.2.2. Rastgele Orman (Random Forest) Algoritması

Rastgele Orman algoritması, 2001 yılında Leo Breiman tarafından geliştirildi. Hızlı çalışan, başarılı çıktılar veren ve yoğun olarak regresyon ve sınıflandırma problemlerinde kullanılan bir algoritmadır [9].

Random Forest, veriyi rastgele parçalara böler. Birden fazla karar ağacını bir araya getirir ve ağaçlar arası rastgelelik ve çeşitlilik oluşturarak tahminde bulunmaya çalışır. Random Forest'in diğer ağaçlardan farkı çok sayıda karar ağacı oluşturmaktır. Bu ağaçları birleştirerek değerlendirme yapabilme yeteneğine sahiptir [10].

3.2.3. Gradyan Artırıcı Ağaç (GBoosted) Algoritması

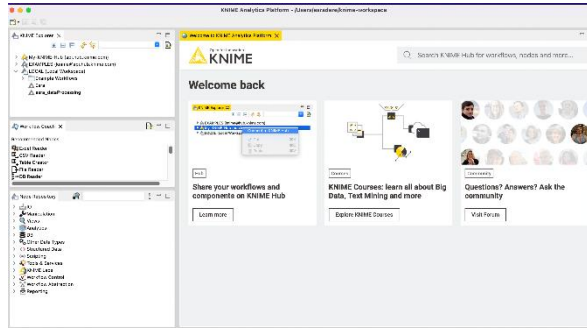
Jerome H. Friedman tarafından bulunan Gradyan Artırıcı (GBoosted) Ağaç, sınıflandırma ve regresyon problemlerinde kullanılan karar ağacı temelli bir algoritmadır. Karmaşık ilişkileri yakalar ve bu ilişkilerle doğru tahminler elde etmeye çalışır. Gradyan Artırıcı Ağaç algoritması ile zayıf tahmin modellerini sıralı olarak birleştirir ve bunlarla güçlü tahmin modellerine dönüştürmeyi hedefler. Her birleştirilen zayıf tahmin ağacı bir önceki ağacın hatasını en aza indirir. Yüksek tahmin performansına sahiptir.

3.2.4. Ağaç Topluluğu (Tree Ensemble)

Ağaç Topluluğu (Tree Ensemble), başarılı bir tahmin modeli oluşturmak için birden fazla karar ağacını birleştirir. Her bir ağaç kendi kararını verir. Sınıflandırma ve regresyon problemlerinde kullanılır. Başarılı sonuçlar elde edilebilen bir makine öğrenmesi yöntemidir.

3.3. KNIME Platformu

KNIME, veriyi analiz etme, görselleştirme, makine öğrenmesi modellemesi, istatistiksel işlevler, tahminleme, kümeleme ve sınıflandırma gibi akışlarımızı yapmamızı sağlayan açık kaynak kodlu bir veri analitiği platformudur. KNIME, veri analiz adımlarının gerçekleştirilebilmesi, verilerin görselleştirilmesi gibi imkanları sağlayan veri keşif ortamıdır [11]. Şekil 5'te KNIME ortamı görülebilir.



Şekil 5. Knime ortamı

Düğümler aracılığıyla işlemler gerçekleştirilir. Bu düğümler analiz sürecini hızlandırır. Eklenen tüm düğümler sırasıyla çalışır ve her düğümün çıktısı görülebilir.

3.4. Veri Seti

Veri seti, Türkiye'deki aktif bir ikinci el araç pazarı olan sahibinden.com sitesinin ilanlarından Selenium test aracıyla 01.03.2023 ile 22.06.2023 tarihleri arasındaki veriler çekilerek MongoDB'ye yazılmıştır. Ham verilerin ön işleme Python kütüphaneleri ile gerçekleştirilmiştir. Veriler aracın fiyatını belirlemede önemli olan yaşı, km bilgisi, modeli, markası, tramer kaydı, kasa tipi, gövde tipi gibi bilgileri içermektedir. Veri seti çekilirken aynı ilanı çekip çekmediği veri alınırken kontrol edilip veritabanına yazılmamaktadır. Fiyat belirlemede etkili olmayan ham veri seti üzerindeki kriterler çıkarılmıştır. Veri setinde eksik değer kontrolü, filtreleme gibi işlemler çalışma kısmında detaylandırılmıştır.

Tablo 1'de gösterilen ham veri setinin bir kısmı gösterilmiştir. Veri seti içerisinde 26.000 üzeri veri ve 28 öznitelik bulunmaktadır. Tabloda aracın markası, modeli, yılı, vitesi tipi, motor gücü, beygir gücü, kimden olduğu, durumu, gövde tipi, kilometre bilgisi, renk gibi bilgiler içermektedir. Ham veri setinde yer alan açıklama, url, açıklama gibi özellikler fiyat tahmininde etken olmadığı için öznitelikler arasından çıkarılmıştır. Veri setinde kullanılan özniteliklerin açıklamaları ve veri tipi Tablo 2'de verilmiştir.

İKİNCİ EL ARAÇ DEĞERLEMESİNDE MAKİNE
ÖĞRENME TEKNİKLERİNİN UYGULANMASI

E. DERE, M. C. KASAPBAŞI

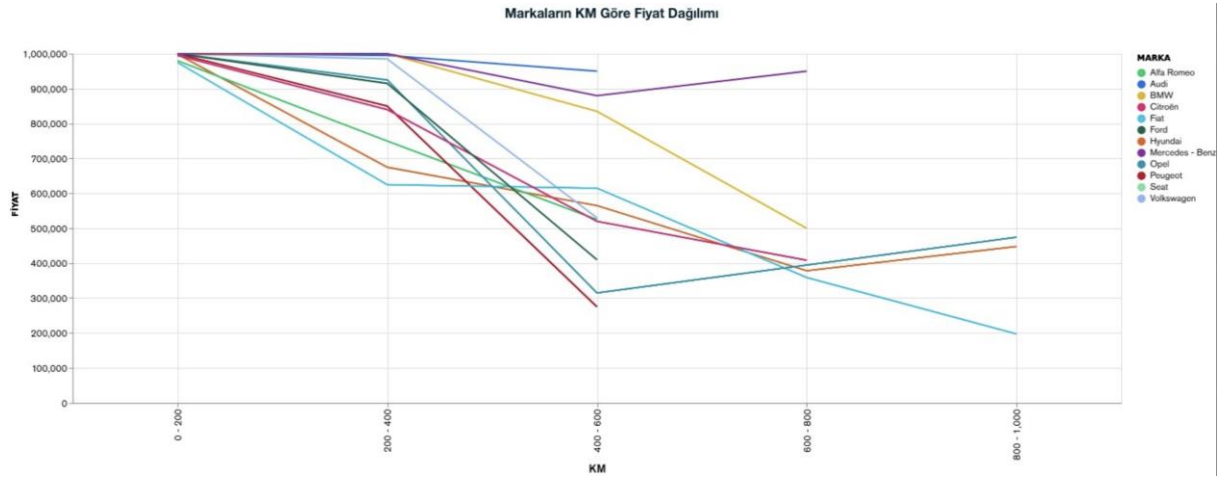
YaratılışTarihi	Marka	Seri	Model	Yıl	YakıtTipi	VitesTürü	Durumu	KM	GovdeTipi	MotorGucu	Çekis	Renk
21 Şubat 2023	Alion	S	580	2021	Elektrik	Otomatik	İkinci El	14500	Sedan	126 - 150 HP	Önden Çekiş	Beyaz
20 Şubat 2023	Alion	S	580	2021	Elektrik	Otomatik	İkinci El	6900	Sedan	126 - 150 HP	Önden Çekiş	Beyaz
15 Şubat 2023	Alion	S	580	2021	Elektrik	Otomatik	İkinci El	6800	Sedan	126 - 150 HP	Önden Çekiş	Beyaz
07 Şubat 2023	Alion	S	580	2021	Elektrik	Otomatik	İkinci El	6500	Sedan	126 - 150 HP	Önden Çekiş	Beyaz
03 Mart 2023	Alfa Romeo	Giulietta	1.4 TB MultiAir Distinctive	2011	Benzin	Manuel	İkinci El	94000	Hatchback 5 kapı	170 hp	Önden Çekiş	Gümüş Gri
21 Şubat 2023	Alfa Romeo	Giulietta	1.6 JTD Distinctive	2014	Dizel	Manuel	İkinci El	94000	Hatchback 5 kapı	105 hp	Önden Çekiş	Kırmızı
09 Mart 2023	Alfa Romeo	Giulietta	1.6 JTD Progression Plus	2014	Dizel	Manuel	İkinci El	154000	Hatchback 5 kapı	105 hp	Önden Çekiş	Beyaz
09 Mart 2023	Alfa Romeo	156	2.0 JTS Distinctive	2004	Benzin & LPG	Yan Otomatik	İkinci El	250000	Sedan	165 hp	Önden Çekiş	Siyah
09 Mart 2023	Alfa Romeo	159	1.9 JTD Distinctive Plus	2010	Dizel	Yan Otomatik	İkinci El	207550	Sedan	150 hp	Önden Çekiş	Gri
08 Mart 2023	Alfa Romeo	147	1.6 TS Distinctive	2006	Benzin & LPG	Manuel	İkinci El	169000	Hatchback 5 kapı	120 hp	Önden Çekiş	Siyah
08 Mart 2023	Alfa Romeo	147	1.6 TS Progression	2004	Benzin & LPG	Manuel	İkinci El	295000	Hatchback 5 kapı	120 hp	Önden Çekiş	Kırmızı
08 Mart 2023	Alfa Romeo	Giulietta	1.6 JTD Progression Plus	2015	Dizel	Manuel	İkinci El	249000	Hatchback 5 kapı	105 hp	Önden Çekiş	Beyaz
08 Mart 2023	Alfa Romeo	147	1.6 TS Distinctive	2006	Benzin & LPG	Manuel	İkinci El	235000	Hatchback 5 kapı	120 hp	Önden Çekiş	Gümüş Gri
08 Mart 2023	Alfa Romeo	146	1.4 TS	1998	Benzin & LPG	Manuel	İkinci El	244000	Hatchback 5 kapı	103 hp	Önden Çekiş	Bordo
08 Mart 2023	Alfa Romeo	Giulietta	1.6 JTD Distinctive	2015	Dizel	Manuel	İkinci El	235000	Hatchback 5 kapı	105 hp	Önden Çekiş	Beyaz
08 Mart 2023	Alfa Romeo	156	1.6 TS Distinctive	2004	Benzin	Manuel	İkinci El	246000	Sedan	120 hp	Önden Çekiş	Mavi
08 Mart 2023	Alfa Romeo	156	2.0 TS Executive	1999	Benzin & LPG	Manuel	İkinci El	241500	Sedan	155 hp	Önden Çekiş	Gümüş Gri
08 Mart 2023	Alfa Romeo	146	1.4 TS Ritmo	1999	Benzin & LPG	Manuel	İkinci El	220000	Hatchback 5 kapı	103 hp	Önden Çekiş	Kırmızı
08 Mart 2023	Alfa Romeo	156	1.6 TS Progression	2004	Benzin & LPG	Manuel	İkinci El	287000	Sedan	120 hp	Önden Çekiş	Füme
08 Mart 2023	Alfa Romeo	156	1.6 TS Distinctive	2005	Benzin & LPG	Manuel	İkinci El	234000	Sedan	120 hp	Önden Çekiş	Füme
08 Mart 2023	Alfa Romeo	Giulietta	1.4 TB MultiAir Distinctive	2013	Benzin	Yan Otomatik	İkinci El	85000	Hatchback 5 kapı	170 hp	Önden Çekiş	Lacvert
08 Mart 2023	Alfa Romeo	Giulietta	1.4 TB MultiAir Distinctive	2012	Benzin	Yan Otomatik	İkinci El	116000	Hatchback 5 kapı	170 hp	Önden Çekiş	Beyaz
08 Mart 2023	Alfa Romeo	Giulietta	1.4 TB MultiAir Distinctive	2011	Benzin	Manuel	İkinci El	79000	Hatchback 5 kapı	170 hp	Önden Çekiş	Siyah
03 Mart 2023	Alfa Romeo	Giulietta	1.4 TB MultiAir Distinctive	2011	Benzin	Manuel	İkinci El	94000	Hatchback 5 kapı	170 hp	Önden Çekiş	Gümüş Gri
21 Şubat 2023	Alfa Romeo	Giulietta	1.6 JTD Distinctive	2014	Dizel	Manuel	İkinci El	94000	Hatchback 5 kapı	105 hp	Önden Çekiş	Kırmızı
09 Mart 2023	Alfa Romeo	Giulietta	1.6 JTD Progression Plus	2014	Dizel	Manuel	İkinci El	154000	Hatchback 5 kapı	105 hp	Önden Çekiş	Beyaz
09 Mart 2023	Alfa Romeo	156	2.0 JTS Distinctive	2004	Benzin & LPG	Yan Otomatik	İkinci El	250000	Sedan	165 hp	Önden Çekiş	Siyah
09 Mart 2023	Alfa Romeo	159	1.9 JTD Distinctive Plus	2010	Dizel	Yan Otomatik	İkinci El	207550	Sedan	150 hp	Önden Çekiş	Gri
09 Mart 2023	Alfa Romeo	159	1.9 JTD Distinctive Plus	2010	Dizel	Yan Otomatik	İkinci El	207550	Sedan	150 hp	Önden Çekiş	Gri
08 Mart 2023	Alfa Romeo	147	1.6 TS Distinctive	2006	Benzin & LPG	Manuel	İkinci El	169000	Hatchback 5 kapı	120 hp	Önden Çekiş	Siyah
08 Mart 2023	Alfa Romeo	147	1.6 TS Progression	2004	Benzin & LPG	Manuel	İkinci El	295000	Hatchback 5 kapı	120 hp	Önden Çekiş	Kırmızı
08 Mart 2023	Alfa Romeo	Giulietta	1.6 JTD Progression Plus	2015	Dizel	Manuel	İkinci El	249000	Hatchback 5 kapı	105 hp	Önden Çekiş	Beyaz
08 Mart 2023	Alfa Romeo	147	1.6 TS Distinctive	2006	Benzin & LPG	Manuel	İkinci El	235000	Hatchback 5 kapı	120 hp	Önden Çekiş	Gümüş Gri
08 Mart 2023	Alfa Romeo	146	1.4 TS	1998	Benzin & LPG	Manuel	İkinci El	244000	Hatchback 5 kapı	103 hp	Önden Çekiş	Bordo
08 Mart 2023	Alfa Romeo	Giulietta	1.6 JTD Distinctive	2015	Dizel	Manuel	İkinci El	235000	Hatchback 5 kapı	105 hp	Önden Çekiş	Beyaz
08 Mart 2023	Alfa Romeo	156	1.6 TS Distinctive	2004	Benzin	Manuel	İkinci El	246000	Sedan	120 hp	Önden Çekiş	Mavi
08 Mart 2023	Alfa Romeo	156	2.0 TS Executive	1999	Benzin & LPG	Manuel	İkinci El	241500	Sedan	155 hp	Önden Çekiş	Gümüş Gri
08 Mart 2023	Alfa Romeo	146	1.4 TS Ritmo	1999	Benzin & LPG	Manuel	İkinci El	220000	Hatchback 5 kapı	103 hp	Önden Çekiş	Kırmızı
08 Mart 2023	Alfa Romeo	156	1.6 TS Progression	2004	Benzin & LPG	Manuel	İkinci El	287000	Sedan	120 hp	Önden Çekiş	Füme
08 Mart 2023	Alfa Romeo	156	1.6 TS Distinctive	2005	Benzin & LPG	Manuel	İkinci El	234000	Sedan	120 hp	Önden Çekiş	Füme
08 Mart 2023	Alfa Romeo	Giulietta	1.4 TB MultiAir Distinctive	2013	Benzin	Yan Otomatik	İkinci El	85000	Hatchback 5 kapı	170 hp	Önden Çekiş	Lacvert
08 Mart 2023	Alfa Romeo	Giulietta	1.4 TB MultiAir Distinctive	2012	Benzin	Yan Otomatik	İkinci El	116000	Hatchback 5 kapı	170 hp	Önden Çekiş	Beyaz
03 Mart 2023	Alfa Romeo	Giulietta	1.4 TB MultiAir Distinctive	2011	Benzin	Manuel	İkinci El	94000	Hatchback 5 kapı	170 hp	Önden Çekiş	Gümüş Gri

Tablo 1. Ham veri seti

Öznelik	Veri Türü	Açıklaması
KayıtTarihi	Object	İlanın Veri Setine Kayıt Tarihi
Kod	Object	İlanın Kodu
YaratılışTarihi	Object	İlanın Oluşturulma Tarihi
Marka	Object	Aracın Markası
Seri	Object	Aracın Model Serisi
Model	Object	Aracın Modeli
Yıl	Int64	Aracın Üretim Yılı
YakıtTipi	Object	Aracın Yakıt Tipi
Vites Tipi	Object	Aracın Vites Tipi
Durumu	Object	Aracın Durumu(İkinci El/Sıfır)
KM	Float64	Aracın kilometresi
KasaTipi	Object	Aracın Kasa Tipi
MotorGucu	Object	Aracın Motor Gücü
Çekis	Object	Aracın Çekişi
Renk	Object	Aracın Rengi
GarantiliMi	Object	Aracın Garanti Durumu
AğırHasarlıMi	Object	Aracın Ağır Hasarlı Olup/Olmadığı
Plaka	Object	Aracın Plaka Uyuşuğu
TakasDurumu	Object	Aracın Takas Durumu
Fiyat	Float64	Aracın Fiyatı
SatıcıTipi	Object	Aracın Kimden Olduğu
Satıcı	Object	İlan Sahibi Bilgileri
Adres	Object	İlanın Yayınlandığı İl, İlçe, Mahallesi
Açıklama	Object	İlanın Açıklaması
Resim	Object	Aracın Resimleri
Url	Object	İlan Linki

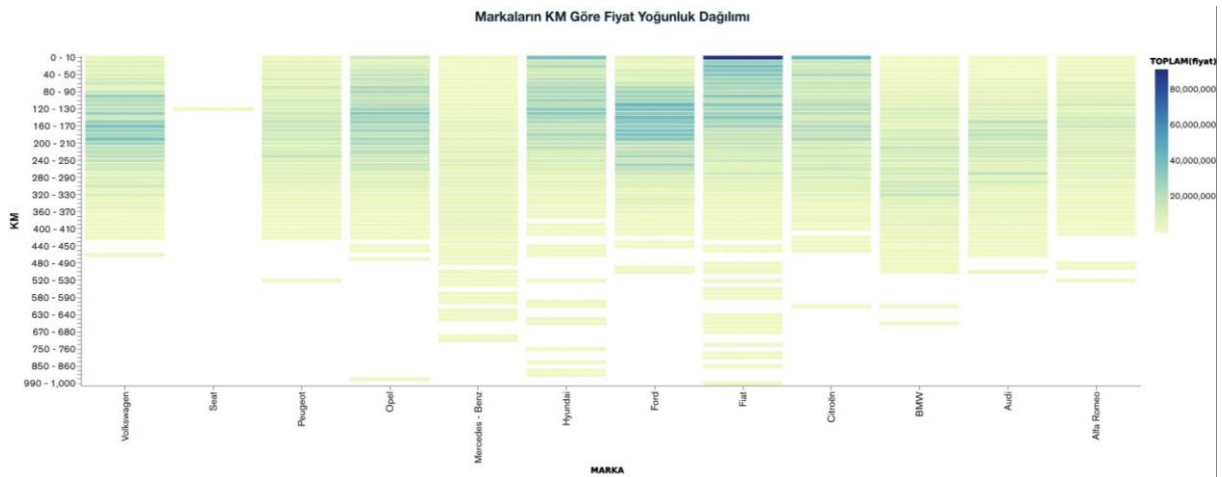
Tablo 2. Veri detayları

Ham veri setinde veri ön işleme adımından önce fiyat hesaplama da etken olan özneliklerle keşifsel veri analizi yapılmıştır. Şekil 6'da markaların km'ye göre dağılımı verilmiştir. Araç markaları incelendiğinde km arttıkça ücretin arttığı gözlemlenmektedir.



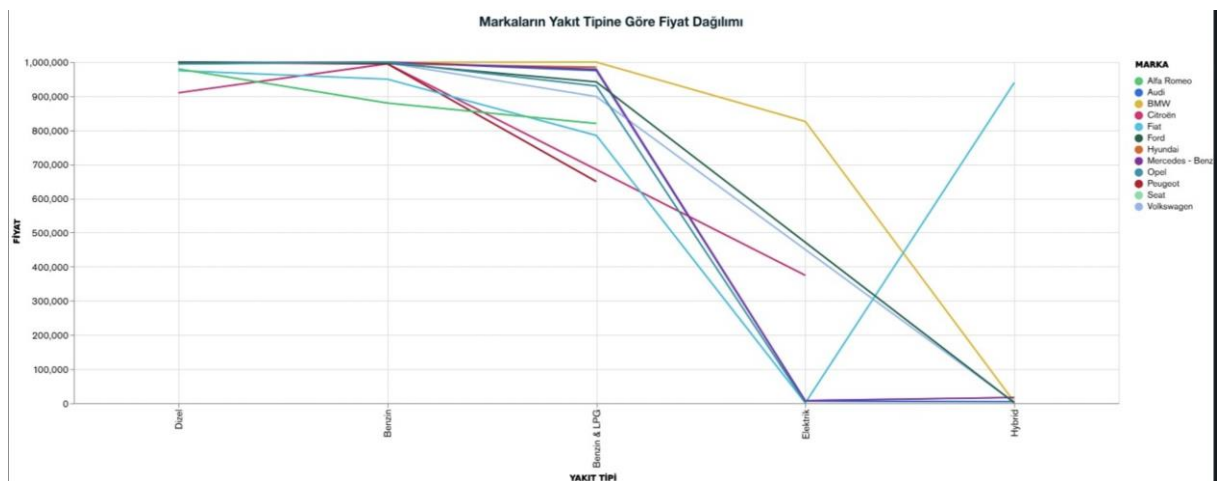
Şekil 6. Km- Fiyat analizi

Şekil 7’de markaların kmye göre yoğunluk dağılımı verilmiştir. Araç markalarının kmsi arttıkça fiyatının arttığı gözlemlenmektedir.



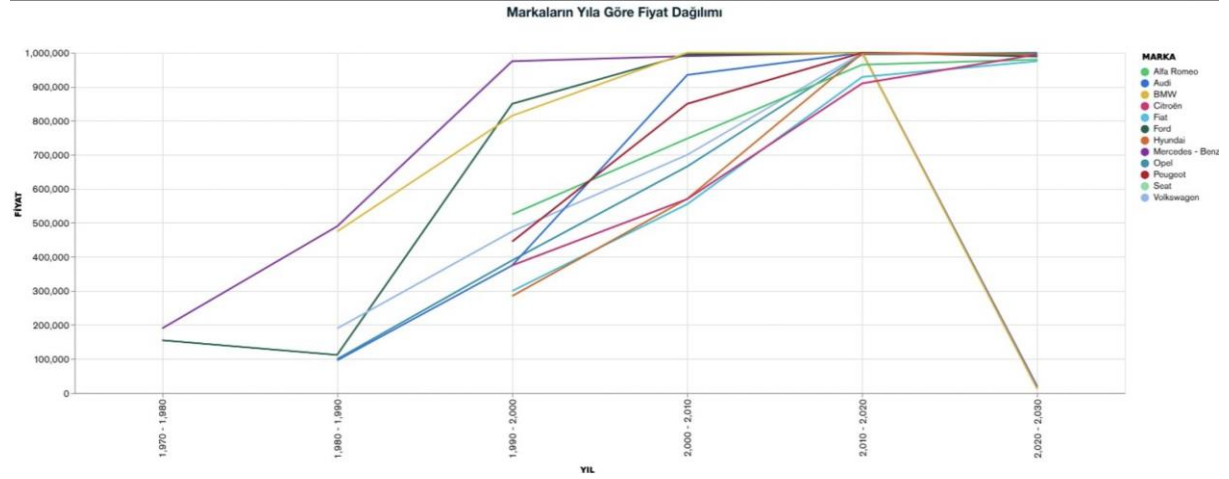
Şekil 7. Marka – Km yoğunluk dağılımı

Şekil 8’de markaların yakıt tipine göre fiyat dağılımı verilmiştir. BMW gibi araç markalarında yakıt tipi benzin olanın fiyatı yüksekken, Fiat araç markasında yakıt tipi elektrikli olan aracın ücretinin daha yüksek olduğu gözlemlenebilir.



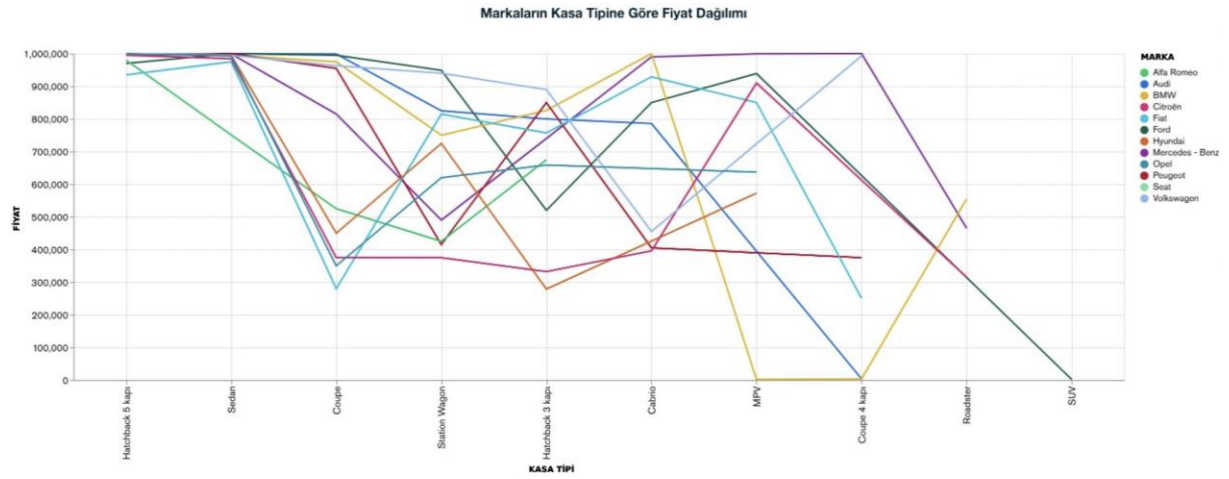
Şekil 8. Marka – Yakıt Tipi fiyat dağılımı

Şekil 9’da yıl arttıkça markaların fiyatlarının arttığı gözlemlenmektedir.



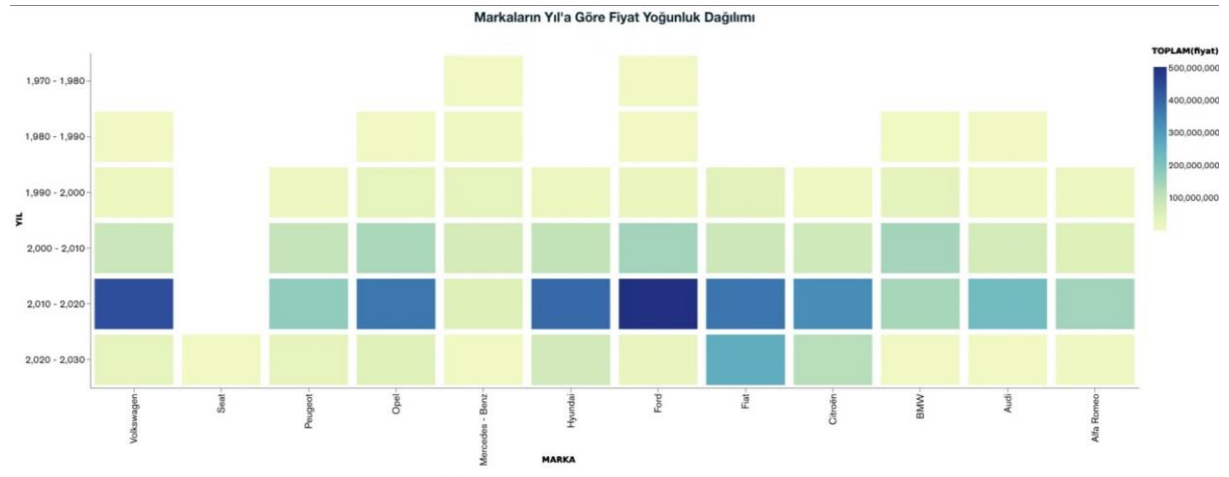
Şekil 9. Markaların yıla göre fiyat dağılımı

Şekil 10'da ise kasa tipine göre araç markalarının fiyatına etki ettiği gözlemlenmektedir.



Şekil 10. Marka – Kasa Tipi dağılımı

Şekil 11'de ise markaların yıl içindeki fiyat yoğunluk dağılımı gözlemlenmektedir.

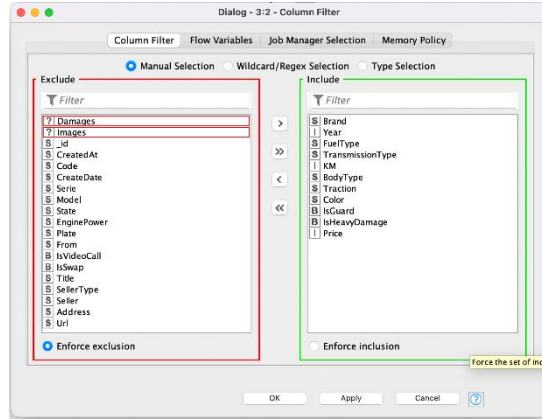


Şekil 11. Marka – Yıl ücret yoğunluk dağılımı

4. Uygulama

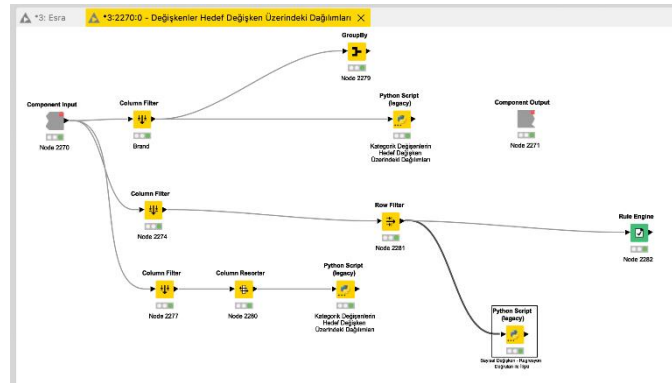
Türkiye'deki e-ticaret sitesinin ilanlarından elde edilen ham veri seti KNIME ortamında veri ön işleme adımlarından geçmiştir. Veri ön işleme adımından sonra aracın değerlemesinde etkili olan bağımsız özniteliklerle bağımlı öznitelik olan hedef değişken fiyat arasındaki ilişkiler incelenmiştir.

KNIME ortamında, excel olarak çıktısı alınmış ham veri seti 'File Reader' düğümü kullanılarak ortama yüklenir. File Reader'a sağ tıklanarak yüklenen veriler gözlemlenebilir. Veri setinde yer alan özniteliklerin hepsi fiyata etki etmediği için fiyat tahminini etkilemeyecek özniteliklerin kaldırılması için 'Column Filter' düğümü uygulanır. Şekil 12'de 'Column Filter' düğümü görülebilir.



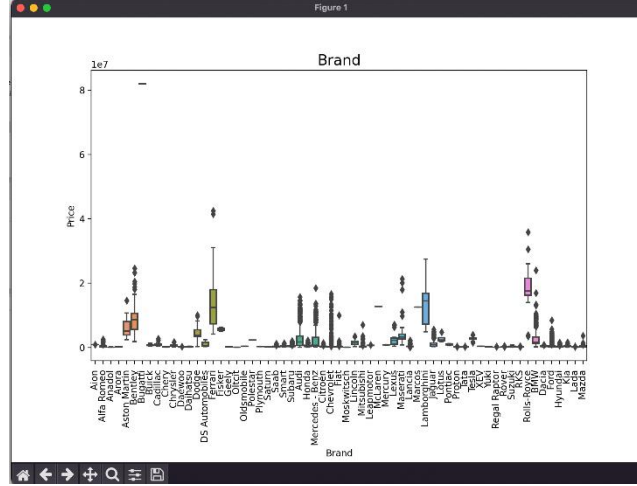
Şekil 12. Sütun Filtresi (Column Filter) düğümü

Sütun filtresi çalıştırdıktan sonra özniteliklerin hedef değişken üzerindeki dağılımı incelemek üzere ve 'Rule Engine' düğümü kullanılarak yeni öznitelik oluşturmak için Şekil 13'te yer alan 'Değişkenler hedef değişken üzerindeki dağılımları' isimli bileşen oluşturulur.



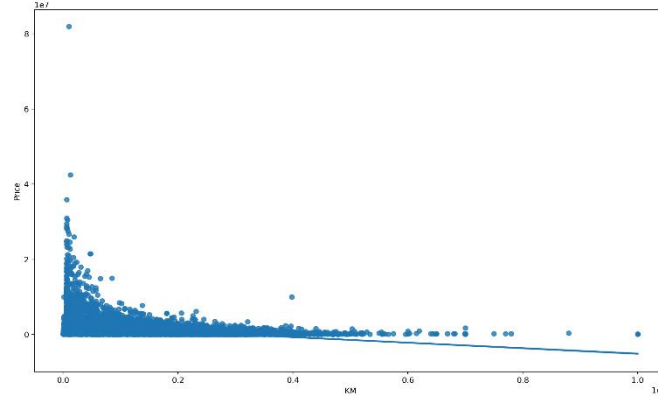
Şekil 13. Değişkenler hedef değişken üzerindeki dağılımları bileşeni

Burada ilk sütun filtresinde marka ve fiyat kolonları alınarak aralarında ilişki Python kütüphanesi olan BoxPlot ile grafikselleştirilmiştir. Şekil 14'te yer alan grafik incelendiğinde fiyatları birbirine yakın araç markalarının gruplanabilir olduğu gözlenmektedir.



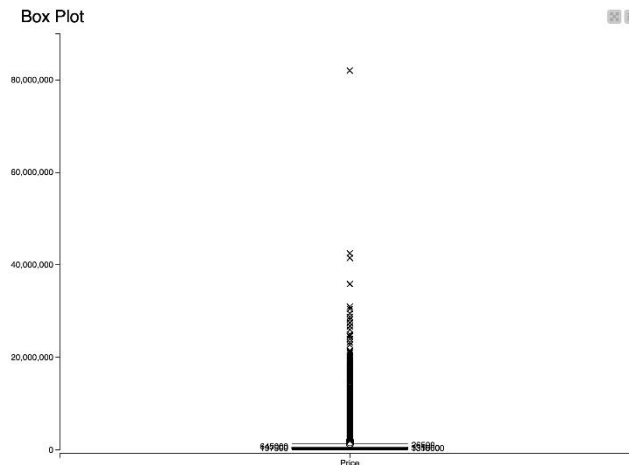
Şekil 14. Marka (Brand) – Fiyat (Price) analizi

İkinci sütun filtresin de km ve fiyat kolonları alınmıştır. Km kolonunda eksik veri kontrolü 'Row Filter' düğümü çalıştırılmış, km-fiyat arasındaki ilişki Python kütüphanesi olan BoxPlot ile grafikleştirilmiştir. Çıkan sonuca göre aracın kmsi artıkça araç değerinin düştüğü Şekil 15'te görüldüğü gibidir.

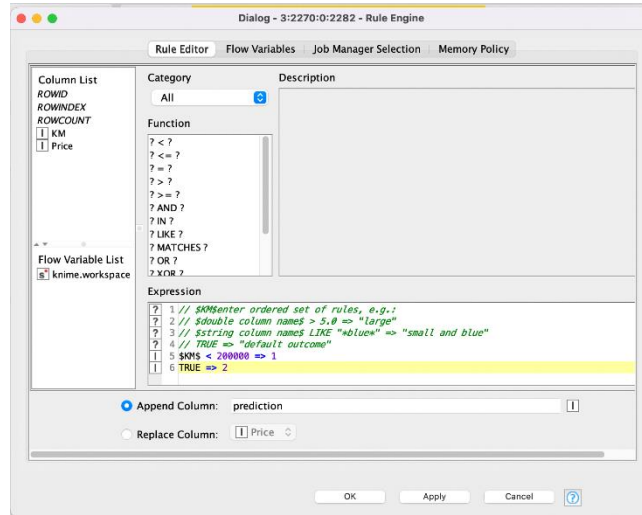


Şekil 15. Km- Fiyat dağılımı

İlk aşamada yer alan 'Column Filter' düğümü adımından sonra 'Box Plot' düğümü bağlanıp çalıştırıldığında veri setinde aykırı değerler(outliers) tespit edilmiştir. Şekil 16'da aykırı değer tespiti görülebilir. Aykırı değerler 'Rule Engine' düğümü ile fiyat tahminine etkisi en aza indirgenmiştir. Şekil 17'de 'Rule Engine' düğümü görülebilir.

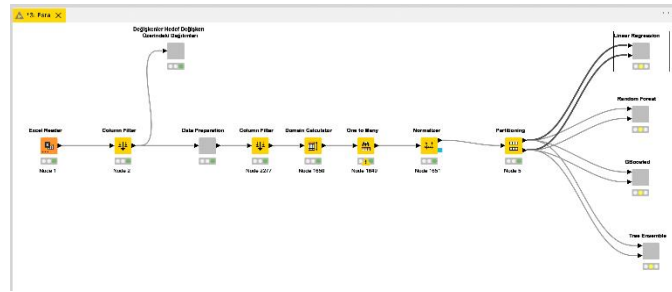


Şekil 16. Aykırı değerlerin tespiti



Şekil 17. Kural Motoru (Rule Engine) düğümü

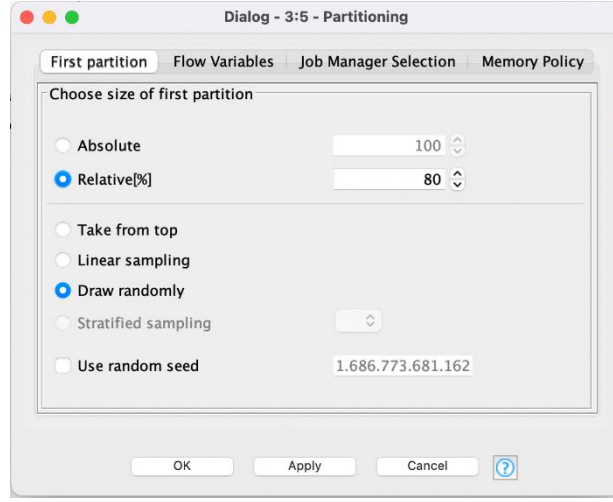
Bu bileşenin çalıştırılmasından sonra Şekil 18’de yer alan düğümler sırasıyla çalıştırılır.



Şekil 18. Uygulama

- Data Preparation düğümü:
Verilerin hazırlanma kısmıdır. Eksik, kirli ve gürültülü verilerin temizlenme aşamasıdır.
- Domain Calculator düğümü:
Bellekteki verileri temizler.
- One to Many düğümü:
Veri setindeki değerleri kategorikten nümerik değerlere çevirir.
- Normalizer düğümü:
Sayının bulunan minimum değer ile farkını alır. Sayının o kolondaki min ve max sayı farkına böler ve normalize eder.
- Partitioning düğümü:
Verilerin eğitim ve test olarak ayrılmasını sağlar.

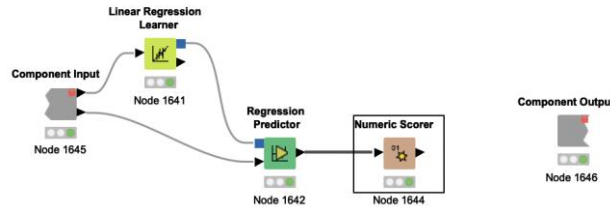
Şekil 19’da veri setinin eğitim ve test olarak ayrıştırılmasına ilişkin görseli verilmiştir.



Şekil 19. Veri setinin eğitim test olarak ayrıştırılması

Veri eğitim ve test dasetine ayrıştırıldıktan sonra metot ve yöntemlerde anlatılan algoritmalar sırasıyla uygulanmıştır.

- Doğrusal (Linear) Regresyon algoritmasının düğümü Şekil 20’de görüldüğü gibidir.



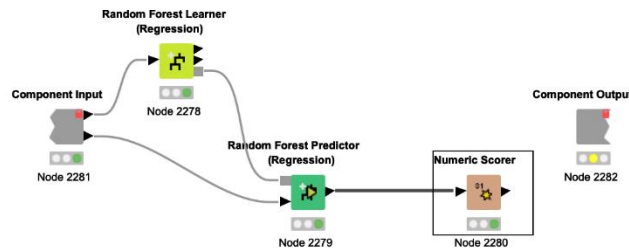
Şekil 20. Doğrusal (Linear) regresyon düğümü

- Tablo 3’te Doğrusal (Linear) regresyonun istatistiksel çıktısı görülebilir.

R^2	0,608
Ortalama Mutlak Hata (MAE)	683.99,02
Ortalama Karesel Hata (MSE)	1.793.888.748.666,89
Kök Ortalama Kare Hatası (RMSE)	1.339.361,32
Ortalama İşaretli Fark (MSD)	-131.103,80
Ortalama Mutlak Yüzde Hatası (MAPE)	1,81
Ayarlanmış (Adjusted)	0,61

Tablo 3. Doğrusal (Linear) Regresyon sonucu

- Rastgele Orman (Random Forest) düğümü Şekil 21’te görüldüğü gibidir.



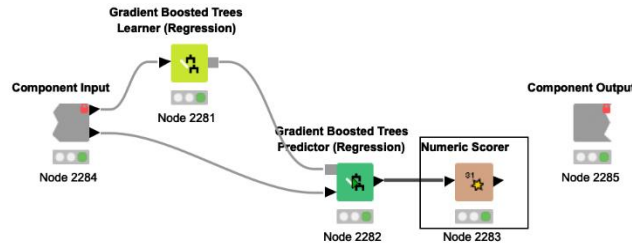
Şekil 21. Rastgele Orman (Random Forest) düğümü

- Tablo 4'te Rastgele Orman algoritmasının istatistiksel çıktısı görülebilir.

R^2	0,85
Ortalama Mutlak Hata (MAE)	255.218,10
Ortalama Karesel Hata (MSE)	686.356.290.885,87
Kök Ortalama Kare Hatası (RMSE)	828.466,23
Ortalama İşaretli Fark (MSD)	6.865,11
Ortalama Mutlak Yüzde Hatası (MAPE)	0,33
Ayarlanmış (Adjusted)	0,85

Tablo 4. Rastgele Orman (Random Forest) sonucu

- Gradyan Artırıcı (GBoosted) algoritmasının detayı Şekil 22'te görüldüğü gibidir.



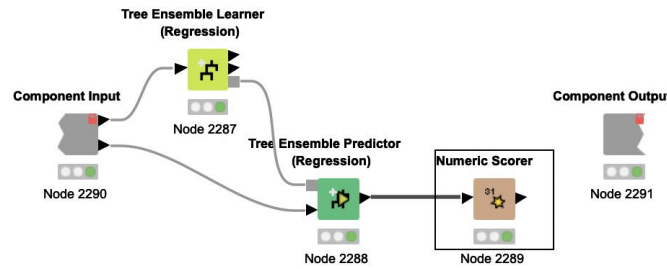
Şekil 22. Gradyan Artırıcı (GBoosted) düğümü

- Tablo 5'te Gradyan Artırıcı (GBoosted) algoritmasının istatistiksel çıktısı görülebilir.

R^2	0,815
Ortalama Mutlak Hata (MAE)	266.626,34
Ortalama Karesel Hata (MSE)	846.837.247.920,42
Kök Ortalama Kare Hatası (RMSE)	920.237,60
Ortalama İşaretli Fark (MSD)	-59.657,90
Ortalama Mutlak Yüzde Hatası (MAPE)	0,29
Ayarlanmış (Adjusted)	0,82

Tablo 5. Gradyan Artırıcı (GBoosted) sonucu

- Ağaç Topluluğu (Tree Ensemble) algoritmasının detayı Şekil 23'te görüldüğü gibidir.



Şekil 23. Ağaç Topluluğu (Tree Ensemble) düğümü

- Tablo 6’da Ağaç Topluluğu (Tree Ensemble) algoritmasının istatistiksel çıktısı görülebilir.

R^2	0,826
Ortalama Mutlak Hata (MAE)	253.209,48
Ortalama Karesel Hata (MSE)	694.143.099.874,02
Kök Ortalama Kare Hatası (RMSE)	833.152,51
Ortalama İşaretli Fark (MSD)	38.585,23
Ortalama Mutlak Yüzde Hatası (MAPE)	0,332
Ayarlanmış (Adjusted)	0,826

Tablo 6. Ağaç Topluluğu (Tree Ensemble) sonucu

5. Sonuç ve Öneri

Bu çalışmada elde edilen ham veri, veri ön işleme adımları ile temizlenmiş ve analiz edilmiştir. İkinci el aracın fiyatını tahmin etmek için makine öğrenmesi yöntemlerinden Doğrusal (Linear) Regresyon, Rastgele Orman (Random Forest), Gradyan Artırıcı (GBoosted), Ağaç Topluluğu (Tree Ensemble) yöntemleri kullanılmıştır. Çalışma, KNIME Analytics Platform veri madenciliği programının 4.7.3 sürümünde uygulanmıştır. Uygulanan algoritmaların başarısını ölçmek için R^2 performans metriği kullanılmıştır. Doğrusal regresyon 0,56 R^2 ile en düşük tahmin oranını vermiş olup, Rastgele Orman (Random Forest) 0,83 R^2 , Gradyan Artırıcı (GBoosted) 0,81 R^2 ve Ağaç Topluluğu (Tree Ensemble) 0,82 R^2 oranları Tablo 7’de görüldüğü gibidir.

Algoritmalar	R^2
Doğrusal (Linear) Regresyon	0,56
Rastgele Orman (Random Forest)	0,83
Gradyan Destekli (GBoosted)	0,81
Ağaç Topluluğu (Tree Ensemble)	0,82

Tablo 7. Tahmin sonuçları

Öneri olarak ikinci el araç satışlarında hem alıcıya hem de satıcıya gerçek fiyatlar sunmayı amaçlayan bu çalışma daha fazla veri setiyle geliştirilebilir. Makine öğrenme yöntemlerinden olan Yapay Sinir Ağları ile tahminleme yapılarak daha başarılı sonuçlar elde edilebilir.

REFERANSLAR

1. Asilkan, Ö.& Irmak, S., “İkinci El Otomobillerin Gelecekteki Fiyatlarının Yapay Sinir Ağları ile Tahmin Edilmesi”, Suleyman Demirel University The Journal of Faculty of Economics and Administrative Sciences, No.2 pp.375-391, 2009
2. Breiman, L, RANDOM FORESTS Leo. 1–33. <https://odayibasi.medium.com/> (17.06.2023)
3. Çelik, Ö. & Osmanoğlu, Ö., “Prediction of The Prices of Second-Hand Cars”, European Journal of Science and Technology, No. 16, pp. 77-83, August 2019
4. Ecer, Fatih, “Türkiye’de 2. El Otomobil Fiyatının Tahmini ve Fiyat Belirleyicilerinin Tahmini”, Anadolu Üniversitesi Sosyal Bilimler Dergisi, 13(4), 101-112, 2013
5. Elshawi R., Sakr S., Talia D., Trunfio P., Big Data Systems Meet Machine Learning Challenges: Towards Big Data Science as a Service, Big Data Research, 14, 1–11, 2018
6. LEE J., Empirical Analysis of Wholesale Used Car Auctions, A Dissertation in University of California, 2006, 120-121, 2001
7. Namlı, E. & Ünlü, R. & Gül, E., “Fiyat Tahminlemede Makine Öğrenmesi Teknikleri ve Doğrusal Regresyon Yöntemlerinin Kıyaslanması; Türkiye’de Satılan İkinci El Araç Fiyatlarının Tahminlenmesine Yönelik Bir Vaka Çalışması”, Konya Mühendislik Bilimleri Dergisi, c.7, s.4, ss. 806-821, 2019
8. ONAT, M. G. Otomotiv Sektöründe Oranlar Yöntemi Aracılığı ile Finansal Analiz, Yüksek Lisans Tezi, Marmara Üniversitesi, İstanbul 2007
9. Özdemir, S., Potential Distribution Modelling and Mapping Using Random Forest Method: an Example of Yukarıgökdere District. Turkish Journal of Forestry | Türkiye Ormancılık Dergisi, (April), 51–56. <https://doi.org/10.18182/tjf.342504>, 2018
10. Pazarlıoğlu, M. V. & Gunes, M., “The Hedonic Price Model for Fusion on Car Market”, International Conference of Information Fusion, Paris, France, 4-13, <http://ieeexplore.ieee.org/document/862707/>, (Erişim:20.08.2016), 2000
11. Tekerek, A., Veri Madenciliği Süreçleri ve Açık Kaynak Kodlu Veri Madenciliği Araçları. Akademik Bilişim’11- XIII. Akademik Bilişim Konferansı Bildirileri, İnönü Üniversitesi, Malatya, ss.161-169, 2011

İNTERNET KAYNAKLARI

<https://www.turhost.com/blog/makine-ogrenmesi-machine-learning-nedir/#serp>, Erişim Tarihi:05.06.2023

Not: Bu makale, İstanbul Ticaret Üniversitesi Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Tezli Yüksek Lisans Programı’nda, Doç. Dr. Mustafa Cem Kasapbaşı danışmanlığında, Esra Dere tarafından yürütülecek olan, “İkinci El Araç Değerlemede Makine Öğrenme Tekniklerinin Uygulanması” başlıklı yüksek lisans tezinin ön çalışmalarından yararlanılarak hazırlanmıştır.